TRAINING EVALUATION IN THE TECHNICAL TRAINING DEPARTMENT AT COMPANY A

by

Rebecca Pearl

A Research Paper

Submitted in Partial Fulfillment of the
Requirements for the
Master of Science Degree in
Training and Development

Approved for Completion of 4 Semester Credits
TRHRD-735 Field Problem in Training and Development

_____

Dr. Joseph Benkowski
Research Advisor

The Graduate School
University of Wisconsin-Stout
May, 2002

The Graduate School
University of Wisconsin-Stout
Menomonie, WI 54751

ABSTRACT

Pearl | Rebecca | L.
(Writer) (Last Name) | (First) | (Initial)

Training Evaluation in the Technical Training Department at Company A
(Title)

Training and Development | Dr. Joseph Benkowski | May/2002 | 50
(Graduate Major) | (Research Advisor) | (Month/Year) | (Pages)

American Psychological Association Society Manual of Style – Fourth Edition
(Name of Style Manual Used in this Study)

There is a tremendous amount of documented research in the field of training evaluation. The Kirkpatrick model still appears to provide the basis for training evaluation in most organizations. In fact, an overwhelming amount of research and publications done in the evaluation field indicate that most organizations use evaluation methods based on Kirkpatrick's four-level model of reaction, learning, behavior, and results. The purpose of this study is to design and implement a training evaluation plan that includes level 2 and level 3 evaluations, along with a written set of standard procedures for evaluation for the Technical Training and Documentation (TTD) department at Company A.

The TTD department does not currently have a comprehensive evaluation process. At this time, TTD only uses level 1 evaluations regularly and level 2 occasionally. These methods, used in this fashion, do not allow in-depth analysis of the training programs. TTD wishes to implement a more thorough evaluation program that includes level 3 evaluations. Level 3 evaluations measure the extent to which the skills learned during training are applied on the job. The results from the evaluations will be used to improve the training courses by identifying deficiencies in the training and highlighting job tasks that the training can target. This study seeks the evaluation methods that are most appropriate for the student population and their work environment. The type of work that the students perform is complex and the tasks are rarely repeated exactly.

Company A utilizes a central training facility and the students travel from around the world to attend the training courses; when the training is complete the students return to their worksites. It is difficult to maintain contact with some of the students due to the fact that they are so geographically dispersed and because many of them work in sensitive environments that do not even have access to email. This makes the follow up contact, which is crucial for level 3 evaluation, extremely arduous. For these reasons, the training department must rely on participation from the students and their managers in the evaluation process.

The study population includes field engineers, site analysts, and district service managers, all of whom are Company A employees. A survey was used to gather data to aid in the development of a level 2 and level 3 evaluation tools. The questions for the survey were derived from the literature review and from working with the manager of the

Technical Training and Documentation department. The results from this survey were analyzed in chapter four. Chapter five contains the conclusions based on the results and analysis of the survey.

TABLE OF CONTENTS

List of Figures

Chapter 1

Introduction

Virtually every industry in corporate America today has some type of training function. There are many kinds of corporate training; the type of training offered by employers to their employees often follows industry trends. During the industrial revolution, factory schools and formal apprenticeship programs provided needed skills for machinists, electricians, and other factory workers (Carnevale, Gainer, and Villet, 1990). During the 1980s corporate training was directed more towards white-collar, college-educated employees (Hawthorne, 1987). Today, corporations provide all kinds of training, including managerial/leadership, new employee orientation, soft skills, sales, and technical.

Many factors are causing an increase in the need for training. Technology is progressing rapidly and a skilled workforce enables businesses to compete in the global market. Product development cycles are becoming shorter, requiring employee retraining more often. Organizational culture is also changing as many workers are no longer lifetime employees, but are changing jobs more often (Brinkerhoff, 1989). And demographics are changing as more baby boomers retire and will need to be replaced with new employees.

All these factors put a burden on training departments. But just as the need for training is increasing, recent years have shown a decline in spending on employee training (ASTD 2001 State of the Industry Report). As this trend continues, training departments must become more accountable for the outcomes of their training programs. Training evaluation, if done properly, can provide a measure of the effectiveness of the training.

Background of the Problem

Company A is the world leader in high-performance supercomputer technology. The company's systems deliver advanced computing to scientific, engineering, academic, and research enterprises. Because of the complex design and packaging, Company A computer systems are maintained and serviced only by Company A field engineers (FEs). There are approximately 140 FEs worldwide. The Technical Training and Documentation (TTD) department is comprised of 19 course developers/instructors, technical writers, editors, and video production personnel. TTD provides all the training and documentation that the FEs require to do their job.

The TTD department does not have a comprehensive evaluation process. It currently uses a simple *Reaction* form that is administered at the end of each class. This does not allow an in-depth analysis of the training programs. TTD has researched methods for gathering further evaluation data, but encountered difficulties. First, once the students leave the training environment, they travel back to their work sites, which are spread throughout the world. So maintaining contact with students and being able to follow their work habits, or *Application* of training, is difficult. Furthermore, many FEs work at secured sites, which makes it even more difficult for TTD to maintain contact with them.

Secondly, the type of work that the FEs perform is complex and the tasks are rarely repeated exactly. One metric commonly used in the computer industry is mean-time-to-repair (MTTR). But there are many factors other than training that impact MTTR, so measuring *Results* of training is almost impossible.

Statement of the Problem

The problem in this study is to design and implement a training evaluation plan that includes level 3 evaluations, along with a written set of standard procedures for evaluation for the Technical Training and Documentation (TTD) department at Company A. The evaluation method must be easily conducted for a student population that is located throughout the world. Standard procedures are important because the majority of the instructors in the TTD department are contractors who work on a part time basis and are located in different locations throughout the U.S. The study will answer the following questions:

- Which evaluation methods are most effective?
- What is the most effective way to perform evaluations with a geographically dispersed student population?

Fisher and Ruffino (1996) recommend establishing goals for training evaluation and identifying the customer of the evaluation. This study establishes an evaluation program in which the training department is the customer of the evaluation and the evaluation goals are to answer the following questions:

- Are the trainees using the newly learned skills on the job?
- If not – why?
- Are there job tasks or skills that were not taught in the training?
- Did the training include information that is not needed?

Significance of the Study

This research is significant to Company A because it is a small and relatively new company, and it doesn't yet have a solid financial structure in place. As Company A faces more competition in the marketplace, customer service becomes a significant factor in sales. Customer satisfaction is often based on a quick, first-time repair of the computer system, so well-trained field

engineers add value to Company A's products. The TTD department is part of the service organization; it doesn't charge field engineers for training, but is financed through service revenue. Because TTD is not a profit & loss center, it must periodically justify its existence. A comprehensive evaluation process would provide real data to help determine the effectiveness and relevance of the training, and more importantly, show areas in which the training and documentation can be improved.

Chapter 2

Review of Literature

Introduction

There is a tremendous amount of documented research in the field of training evaluation. Evaluation is generally defined as a judgement process that determines the effectiveness of a program. According to Kirkpatrick (1994), some define evaluation as measuring changes in behavior that results from training, while others feel that real evaluation determines the overall results of the training program. While there are many reasons to evaluate training, an increased emphasis on accountability of training and human resource development (HRD) departments has increased awareness of evaluation methods (Phillips, 1997).

The Kirkpatrick model still appears to provide the basis for training evaluation in most organizations. In fact, an overwhelming amount of research and publications done in the evaluation field indicate that most organizations use evaluation methods based on Kirkpatrick's four-level model of reaction, learning, behavior, and results. But in the 40 years since Kirkpatrick published his method, many training organizations have found limitations in the model and have modified the four levels to meet their needs. This chapter documents the research on evaluation methods that are used by training organizations; problems often encountered in evaluation and their subsequent solutions; and the methods that are most often employed to provide meaningful evaluation data.

Level 1 Evaluation – Reaction

Kirkpatrick's level 1 is referred to as student *reaction*. This simply indicates how well the participants liked the training program. Some of the data sought in this type of evaluation include feedback about the instructor, the delivery method, course content, and

the training facility. In order for the reaction level evaluation to be effective, Kirkpatrick (1994) suggests that the evaluator design the evaluation tool as follows:

- Determine what you want to find out
- Use a form that can quantify results
- Make the form anonymous
- Allow space for additional written comments

A level 1 evaluation usually takes the form of a survey that the training participants complete at the end of the training, during what is referred to as the "post-training event euphoria" in a study by Mathews et al. (2001). Although level 1 is the most commonly used form of training evaluation, it is often considered the weakest form of evaluation because a positive reaction cannot predict a training program's success. One reason for this weakness is that the training participants may not be the best judges of the training's effectiveness. Most of the data is based on how well the participants liked the trainer (Hubbard, 2001). McCurry (1999) concurs that the end-of-course reaction does not indicate how well the knowledge will transfer to the job. Kidder and Roullier's (1997) work supports this finding; they report that "Favorable reactions to training do not guarantee that trainees have learned anything or that they will use what they have learned back on the job." Use of level 1 evaluation often causes trainers to put an emphasis on making the training enjoyable, but not necessarily effective.

The evaluation tool itself can present another problem with level 1 evaluation. Hubbard (2001) feels that many level 1 evaluations are poorly designed, contain vague questions, and do not mirror the learning objectives. Although this problem should be relatively easy to correct, many trainers who administer these evaluations do not have the skills or resources to write them effectively or to study level 1 results in depth (Bernthal, 1995).

Potential problems can arise when analyzing results from a level 1 evaluation. Trainers often do not compare evaluation results among different groups of training participants. Level of expertise and job responsibility can cause differences in results. Trainers can make incorrect conclusions when evaluations are analyzed individually (Brinkerhoff, 1989). A trainee's learning style can affect his or her reaction to the training and can be a factor in learning as well. Lippert, Radhakrishna, Plank, and Mitchell (2001) note the use of learning style inventories to determine the learning style of the training participants. In their study, they were particularly interested in how learning styles impacted the outcomes of web-based training.

Although level 1 evaluation doesn't measure learning, it can aid in determining the effectiveness of a training program. A trainee's reaction is important because if they are not satisfied with the training during the training event, they will not be motivated to learn. Kirkpatrick (1994) states that the training participant must like the training program to reap the maximum benefit from it. Level 1 evaluation can gauge how well the participants accepted the training program. Kidder and Rouille (1997) discuss reasons why positive reaction to training is important: they feel that positive reaction leads to organizational support of the training, and students who feel positively about the training are more likely to apply the new skills on the job.

Many trainers use modified methods for performing level 1 evaluation to gain more meaningful data. In a study of manager safety training programs conducted by the Environmental, Safety and Health training team (Haberstok, 1997), level 1 evaluations used open-ended questions. The instructors also explained to the training participants who their feedback is used by and why it is important. Level 1 evaluation can be thought of as a measure of customer satisfaction. Dixon (1996), in a study of more than 40

organizations, found that the companies with the best practices put the focus of level 1 evaluation on the customer, rather than the trainer. Although this method could be thought of as needs assessment, her study found that at IBM, over half of what the company identified as customer requirements of the training were not measured in any evaluation. When customer requirements are identified, they can be turned into the standards that are measured by an evaluation tool.

Lawson (1998) cautions against designing the level 1 evaluation tool to measure things you cannot control. She recommends determining exactly what you want to measure. Some of the items she does recommend including in the evaluation are:

- Content
- Materials
- Instructional method
- Trainer
- Environment
- Logistics

Lawson, like Haberstok (1997), also advocates giving the respondents an opportunity to make suggestions to improve the training. Most evaluation researchers recommend using a multiple-response format in the evaluation tool. This prevents a respondent from simply circling the same number in a Lickert scale, and forces them to give thoughtful feedback. Lippert (2001) suggests modifying the traditional level 1 evaluation form to gather more information. His level 1 evaluation includes a student's self-assessment by asking what they think they learned after the training.

Kirkpatrick's experience shows that a training director should also make an appraisal of the training program, and that appraisal should be used in combination with the training participants' feedback from the level 1 evaluation.

Level 2 Evaluation – Learning

Kirkpatrick's level 2 evaluation measures the *learning* that takes place during the training. Although there are many definitions of learning, Kirkpatrick defines it as the principles, facts, and techniques that the trainees understood and absorbed (1994). Friedman and Yarbrough (1985) add to that definition the knowledge of how, when, and where to appropriately apply the newly learned knowledge.

During level 2 evaluation, the learner typically demonstrates the new skill or is tested on the knowledge. Measuring learning does not have to be a difficult process. Student demonstration of the skill is the most widely used method. Kirkpatrick (1994) suggests comparing the performance of a control group that has not received the training and the group that has been trained. This method provides data to correlate performance to the training. But for this level of evaluation to be truly effective, it must be based on the training objectives. Both Kirkpatrick and Lawson (1998) stress the importance of measuring the learning based on the course objectives. Bee (2000), agrees that well-formed objectives are the foundation of evaluation.

Level 2 evaluations are sometimes referred to as objectives-based studies. This type of evaluation works well in criterion or skills based training. And the use of the training objectives can aid in organizing the evaluation. Criterion-referenced tests and samples of students' work are used together to assess the outcome of the training (Stufflebeam, 2001).

Lawson (1998) feels that the most appropriate methods for measuring learning include testing and observation. Testing is probably the most frequently used method. There are many styles of tests, but in skills-based training, objectives tests are most appropriate. Lawson notes that some of the advantages of objective testing are that they

are easy to score and inexpensive to use. Observation usually consists of watching the student demonstrate of the skill.

Another method for measuring learning is the use of pre- and post-training tests. Lippert, et al. (2001) used pre- and post-testing to measure learning for Internet-delivered inservice training conducted by county Extension agents. Internet delivery was a new method for this type of training, and the authors of the study were particularly interested in how the delivery method affected learning. Pre- and post-testing doesn't have to involve traditional paper-and-pencil tests. There are many other effective methods for this type of testing.

Birkland (2001) relates how the regional manager of a trucking fleet uses pre- and post-testing in his technician training programs. Many learners can feel intimidated by the testing process, so the trainer should try to make it fun and non-threatening. For the pre-test, the students rate themselves on their ability to train the other members of the group by using the same training materials. The process is repeated at the end of the training. The results usually indicate an increase in the learners' confidence level and help them realize their increased knowledge. Mack Trucks, Inc. uses pre- and post-tests not only for the learners' benefit, but also to evaluate the training materials and the evaluation tools themselves. For example, if one post-test item is consistently answered incorrectly, they will look at the wording of the test question and at the training materials that address the missed test question (Birkland, 2001).

Dixon (1996) found that level 2 evaluation by use of testing serves two purposes at Motorola. The test results provide information for course improvement during the pilot phase of course development. Test results are also used to provide feedback to training participants about their learning. The participants can receive remedial help if their test

score is below a certain level. Motorola embeds the testing during the course, rather than waiting until the end of training. Dixon (1996) also studied FPL Nuclear's use of testing. They use pre-testing to determine the level of training that is needed. Post-testing certifies that the learners are qualified to perform critical tasks. FPL Nuclear also tests participants several months after the training to determine the need for refresher training.

Another way to perform level 2 evaluation is by using in-training evaluation (ITE) programs. These programs were documented by Turnball, Gray, and MacFadyen (1998), all medical doctors, who use this method in the Department of Medicine at the University of Ottawa. The authors of this article suggest the use of written exams, objective-structured clinical examinations, observations, formal oral examinations, and chart-stimulated recall assessment. These methods, which focus on performance assessment, were chosen because traditional assessment methods failed to provide sufficient information to truly evaluate the level of learning among the students. While most of the skills they evaluate are technical in nature, observation methods were found to be particularly useful in assessing communication skills, ethical behaviors, reliability, and integrity. The authors of this article did point out some of the limitations with the ITE method. One of the problems that can be found in most training evaluations is that the teacher is usually the evaluator and he or she may not be properly trained to carry out the evaluations. Another disadvantage is that observations are time consuming, and when they are used, they may not be documented frequently enough to provide valid data. Eight principles should be considered in order to develop a meaningful ITE program: reliability, validity, flexibility, comprehensiveness, feasibility, timeliness, accountability, and relevance. This sound ITE program helps the medical college identify students who

experience difficulty and encourages learning to the objectives of the training program (Turnbull, et al., 1998).

Like level 1 evaluation, level 2 has its limitations, especially when used in the absence of other evaluation methods. A student may be able to demonstrate at the conclusion of training that he or she has learned a skill, but that does not guarantee that the skill or knowledge will transfer to on-the-job behavior (Kirkpatrick, 1994). This is why level 3 evaluation is important.

Level 3 Evaluation – Application

Level 3 evaluation is known as *application* or transfer of knowledge. In business and industry this typically refers to the use of newly learned skills and practices on the job. Blanchard, Thacker, and Way (2000) document research that shows that only 62% of US companies with 100 or more employees evaluate at level 3. Kirkpatrick (1994) points out three decisions that must be made when designing level 3 evaluations: when is the best time to evaluate, how often to evaluate, and which evaluation method to use.

Although many researchers find that level 3 evaluation is time consuming and costly, it does provide relevant information regarding the lasting effects of training. Level 3 evaluation is important because even though level 2 determines whether the trainee mastered the objectives of the training, there is no guarantee that the skills will transfer to work behaviors on the job. In fact, Hubbard (2001) suggests that a person should be considered trained when he or she can demonstrate the training objectives. Whether or not the trainee uses the skill in the workplace is not the responsibility of the trainer. It is difficult to measure just how much training affects job behavior because many factors affect behavior on the job, and training may be only a small factor. Kirkpatrick (1994) lists five requirements for change in job behavior:

- Desire to improve performance
- Recognition of one's own weaknesses
- Permissive work environment
- Help from an interested and skilled individual
- Opportunity to use the newly learned practices or ideas

Assessing job behavior as a result of learning is easier in production jobs than in white collar or knowledge jobs. Production job behavior can usually be measured in quantifiable terms such as increased or decreased output. However, the job duties of managerial workers do not always produce tangible results and do not fall into easily measured classifications (Hawthorne, 1987).

Dixon (1996) found that at the companies she studied, level 3 evaluation is used selectively, but not consistently. She identifies three primary concerns about this trend. The first is that no standard form exists that can be used to gather this type of data. The data collection method must be customized for each situation. Second, most trainers in these organizations feel, as most do, that evaluation at this level is time consuming and costly. Lastly, level 3 data collection requires collaboration between clients. Even though Dixon's study of Arthur Andersen found that they use level 3 evaluation in fewer than 10% of training programs (it was used when the data was useful to the client), the results from these evaluations greatly enhanced the credibility of the training. Dixon also discovered that none of the organizations used level 3 results for the purpose of justifying training programs or budgets.

Although many researchers and authors find much disagreement regarding which methods work best, most researchers do agree on general methods for level 3 evaluations. For example, Kirkpatrick (1994), Laird (1978), Phillips (1997), and May, Moore, and Zammit (1990) all agree on using surveys/interviews, observations, and testing as valid methods. Lawson (1998) also recommends the use of observations, interviews, and

surveys to conduct level 3 evaluations. During an observation, the trainer or another qualified observer watches the employee perform their job. The observer should use a checklist or some other data collection form to ensure consistent results. While this method provides a firsthand look at job behavior, it is costly to implement and the employee may resent this type of evaluation. For interviews, Lawson advises interviewing the training participant's manager, coworkers, customers, and subordinates, in addition to the training participant. Mack Truck, Inc. uses this method when the trainers make follow-up calls to the Mack dealer or service managers to inquire if the managers see a difference in the work behavior of the employees who were trained (Birkland, 2001). Surveys may be less expensive to use than interviews and observations. As with interviews, Lawson (1998) suggests surveying not only the training participant, but others associated with the participant as well.

Having clear expectations for the training is the key to ensuring lasting, positive effects on the job. Wade (1994) advocates the following process for defining the job actions that should result from the training:

- Develop action objectives
- Gather input from those directly involved with the action
- Determine how and when to survey those involved
- Use written assessments by the trainees themselves

Post-training assessments need to be conducted at regular intervals. Wade recommends performing the evaluations at 3 months and 6 months after the training event. Using regular long-term assessments allows you to determine the lasting effects of the training. Fisher and Ruffino (1996) agree that the first step in level 3 evaluation is to identify the job performance data to be measured. They recommend using training objectives, subject matter experts, and performance standards specific to the trainee's job functions. Like other researchers, Fisher and Ruffino recommend gathering data from

training participants, their managers, etc., but also surveying internal and external customers and QA /audit personnel. They also list documentation, such as performance tracking reports, error reports, attendance records, complaint records, and work samples as sources of applicable data.

Deming (1982) documented a process developed by the US Office of Personnel Management, known as Participant Action Plan Approach (PAPA). This process advocates requiring the trainee to create an action plan that they take back to their job after the training. The plan is developed at the end of the training event and lists behaviors the trainee wishes to apply on the job. After a predetermined time, the trainee is contacted to see what new job behaviors they have implemented. Once the data is collected, it can be summarized numerically, qualitatively, or by using a combination of quantitative/qualitative methods. Deming then suggests using the collected data to make recommendations, if applicable. Like other level 3 methods, Deming's method requires proper planning up front and the support of training participants and management.

One important aspect of Deming's (1982) model that differs from some others is that he also asks trainees about any problems in the work environment that hinder application of new job behaviors. Similarly, Kidder and Rouiller (1997) advocate gathering data regarding environmental factors that *promote* the application of skills. This is an area that is often ignored by training evaluators. Abernathy (1999) writes that training evaluation should deal with psychological and sociological issues and not just business results that can be assigned a monetary value. In a study at General Electric, she found that their philosophy is that training results come in many forms and that intangible results, such as improved customer service and career development for GE employees, are as valuable as tangible results. At GE a method was formulated to measure factors

that promote or inhibit the use of training skills. The company used a survey that asked questions regarding different aspects of the work environment. Respondents rated each survey item on a scale of 1-never to 9-always. This numerical rating made it easy to apply statistical analysis to the data.

Most evaluators use a standard set of methods to conduct level 3 evaluation, but there are other ways to identify changes in work behaviors. Blanchard, et al. (2000), found that some companies who indicate that they do not perform level 3 evaluations actually *do* gather this type of data, but not by using traditional methods. For example, some organizations use "integrated human resource systems" in which training and other human resource practices work together to focus on the organization's strategic plan. In this method, a performance review process assesses on-the-job behaviors and the organization's effectiveness together, not as isolated factors. Training is not viewed as a separate method to assist employees to do a better job, but rather as part of the work process itself. So even though organizations that use a strategic human resource management model do not focus on training evaluation alone, they do evaluate their entire HR system, which includes training.

One of the reasons most often cited as to why skills are not transferred on the job is a lack of management support. Bernthal (1995) mentions lack of management support as a factor that often undermines an effective training program. Phillips (1997) calls management support "critical" to application of new skills on the job. In fact, he advocates management involvement throughout the entire training process. This involvement starts with pretraining program support. The training department and the trainee's supervisor should write an agreement that defines what is expected from the training program as well from the supervisor after the training has taken place. Managers

should also encourage training programs for their employees and offer assistance, if appropriate. One of the key actions by management is support and reinforcement of the behavior changes that result from training. Without this reinforcement, the use of newly acquired skills often declines. Phillips also recommends that managers attend the training program first. By experiencing the program firsthand, they will be familiar with the course content and will have a clear understanding of the program. Another method to ensure reinforcement is through management reinforcement workshops. These programs specifically address how to reinforce work behaviors that are taught during the training programs.

Level 4 Evaluation – Results

Level 4 evaluation measures the *results* of training, usually in terms of financial impact. When stated in simple terms – determining whether the desired results of training occurred – level 4 evaluation appears to be easy to accomplish. But this is usually the most difficult phase of evaluation to perform. Kirkpatrick (1994) himself states that in some work environments, it is "extremely difficult if not impossible" to evaluate at this level. Hubbard (2002) agrees strongly: "In almost all cases, the financial impact of a training event cannot be measured." Most researchers point to the many variables that affect work performance as the main hindrance in achieving reliable results data.

Phillips (1997) defines this problem as a critical issue. The first step in conducting level 4 evaluation is isolating the effects of training from other variables that affect job performance. He found that in some work environments, level 3 and level 4 evaluations are not even planned because the data is simply not available. This lack of data usually occurs in job functions that are not technical or skills based, such as customer service jobs, or work environments in which it is difficult to attach a monetary value to quality

improvements. Dixon (1996) agrees that it is impossible to prove that work results are the sole effect of a training event. However, she found that the companies she studied were more concerned with whether they obtained results, than with proving that training was the cause of the desired results.

Most trainers assume that level 4 evaluation is more valuable than other evaluation methods and provides more meaningful data. But this may not be true for all training programs especially in the area of interpersonal skills training. Although quantifiable measures are typically used in level 4 evaluation, this type of data is just not as relevant for interpersonal skills training as level 1 or level 2 data might be (Bernthal, 1995).

Phillips (1997) agrees that not all evaluation data can be measured in financial terms. For some training, the intangible results are just as important in the evaluation process. Kaplan and Norton (1992) wrote, "financial measurements worked fine in the industrial era, but now they're out of sync with the skills and competencies that companies measure."

Return On Investment (ROI)

Another type of training evaluation, beyond Kirkpatrick's four levels, is measuring return on investment, or ROI. This type of evaluation is similar to Kirkpatrick's level four in that it seeks to identify training results, usually in a monetary form. But it differs because it also examines the cost of the training (investment) that may have led to the measured results. The cost of training is subtracted from the results to identify the ROI. Interest in ROI appears to be growing; articles on ROI appear more frequently in training and HRD publications and there are even books devoted to the

subject of ROI. There are many reasons for the increased interest; Phillips (1997) cites competitive economic pressures as one cause of closer scrutiny of training costs.

Measuring ROI is a controversial subject in many training and HR departments. Some trainers do not pursue measuring ROI because they feel it is simply not possible. Many training departments are not set up as profit and loss centers, so there is no need to measure ROI. Others are required to produce ROI data in order to justify their budgets. The consensus among most researchers and publishers on this subject is that measuring ROI is not appropriate for every type of training program. Whether or not companies perform ROI studies, most practitioners agree that it is difficult to accurately measure ROI.

Research points to many reasons for the difficulty in measuring ROI. ROI studies are often costly and time consuming, and most training evaluators are not knowledgeable about ROI. One of the most common concerns is that although it may be easy to calculate the costs of training in financial terms, many training outcomes cannot be measured this way. Phillips (1996) defines such training outcomes as "soft data" and they include items like employee absenteeism, job satisfaction, employee loyalty, employees' perceived changes in performance, problems solved, and frequency in using new skills. And, as with level 4 evaluation, training is only one of many factors that influences work productivity, so it may be unfair to hold training accountable for work results.

Another hurdle often identified is the resources required to perform a thorough ROI evaluation. Goldwasser (2001) cites Verizon's Workforce Development Division spending about $5000 for a single ROI impact study. It is easy to see how this could strain the budget of many smaller training or HR departments. Parry (1996) reinforces this strain on resources by stating that many trainers lack the accounting skills and the

time to do this type of analysis. He also cites the fact that training benefits often accrue over time. When this is true, how does an organization know when to measure the training outcomes? But one of the biggest fears about analyzing ROI is that the results may be negative and may harm the training department and impact budgets. Phillips (1997) also mentions fear as a factor in not performing ROI evaluations. There are several reasons for this fear: fear of the unknown, consequences of a negative ROI, and fear that the results of the ROI analysis may cause changes in the organization.

Even with all the negative aspects associated with ROI evaluation, there are valid reasons for doing it. In the big picture, the overall objective of business is to make a profit, and training is one function that should contribute to that goal. Many training professionals prefer financial data to justify their training programs. And when ROI analysis does indicates that training has contributed to the bottom line, it is easier to justify the cost of training programs (Goldwasser, 2001). Another advantage of performing ROI analysis is that when trainees and their managers take part in the ROI process, they tend to be more committed to the training. Managers will also be more involved in the trainees' development plans. The advantage to the trainer is that the course will be more relevant and focused on financial results, when it is appropriate. Training management can better determine any weakness in the training so it can be revised and improved (Parry 1996).

So just how are ROI studies done? One area in which it would seem to be easy to measure ROI is sales training because in the sales function, improved productivity – increased sales – should be easy to measure in monetary terms. But surprisingly, research on sales training ROI indicates that companies do not use simple mathematical calculation to measure ROI. Beveridge (2000) advises that the first step in evaluating at

the ROI level training is to set expectations and requirements for increased productivity prior to sales training. He offers specific examples, such as increased sales, improved margins, and new sales accounts. Sales managers should meet with the sales people before the training event and help them plan what the improvements and new skills that will be addressed in the training, and then set a date at which those results will occur on the job. Management must also hold the salesperson accountable for those results. This method puts decreased emphasis on dollars and cents, encourages management involvement, and focuses on the job behaviors of the salesperson, which will eventually result in increased revenue.

Keenan (2000) looked at sales training ROI in an examination of practices at several major U.S. corporations. Like Beveridge (2000), Keenan's study also found more of an emphasis on performance than strictly on financial figures. In fact, he found that even Xerox Corp. "treads cautiously around claims of dollars-and-cents return on its investment." They ask the training participant to assess whether an increase in job productivity resulted from a specific training event. While Xerox intends to analyze the relationship between training investment and results in quantitative terms, they agree that it is difficult. Keenan also found that Motorola Inc. uses a method of integrating training with work results and comparing those results closely to the training event. Their training is focused on business-case-based proposals. Metrics are used to assess job performance. Examples include measuring the number of prospects that the salesperson contacted and the number of sales proposals generated. While Motorola management concedes that you can't put a dollar figure on proposals or contacts, if sales rates do go up, revenue increases. Like Xerox, Motorola also involves sales management in supporting salespeople throughout and after the training event. This incorporates training into

everyday job activity, at which point training results are indistinguishable from sales results. This de-emphasizes the need to measure ROI.

Summary

The review of literature shows that most training organizations employ a variety of evaluation methods and those methods relate to Kirkpatrick's 4-level method of reaction, learning, application, and results. Although each evaluation level has its limitations, they can be modified to work with just about any training/workplace environment.

Measuring ROI is controversial and not widely used, but interest in ROI is growing. Gathering valid data is not easy because it is hard to isolate the effects of training from other factors that influence financial results.

Chapter 3

Methodology

Introduction

The Technical Training and Documentation (TTD) department does not currently have a comprehensive evaluation process. At this time, TTD only uses level 1 evaluations regularly and level 2 occasionally. These methods, used in this fashion, do not allow in-depth analysis of the training programs. TTD wishes to implement a more thorough evaluation program that includes level 3 evaluations. Level 3 evaluations measure the extent to which the skills learned during training are applied on the job. The results from the evaluations will be used to improve the training courses by identifying deficiencies in the training and highlighting job tasks that the training can target.

There are different methods available to conduct this type of evaluation. The review of literature revealed that the most effective methods for level 3 evaluation are:

- Surveys
- Interviews
- Testing
- Follow up phone calls
- Self assessments
- Gathering data from:
  - Performance tracking reports
  - Error reports
  - Compliant records
  - Work samples

This study seeks the evaluation methods that are most appropriate for the student population and their work environment. The type of work that the students perform is complex and the tasks are rarely repeated exactly. Isolating the effects of training is difficult because there are so many other factors that can impact job performance.

Company A utilizes a central training facility and the students travel from around the world to attend the training courses; when the training is complete the students return

to their worksites. It is difficult to maintain contact with some of the students due to the fact that they are so geographically dispersed and because many of them work in sensitive environments that do not even have access to email. This makes the follow up contact, which is crucial for level 3 evaluation, extremely arduous. For these reasons, the training department must rely on participation from the students and their managers in the evaluation process.

Study Population

The study population includes Field Engineers (FEs), Site Analysts (SAs), and District Service Managers (DSMs), all of whom are Company A employees. There are approximately 120 FEs, 30 SAs, and 17 DSMs worldwide. The FEs job duties include maintenance and repair of computer systems. When a customer purchases a computer system from Company A, it must also purchase a maintenance contract. This contract allows only Company A employed and trained FEs to service the computer system. If the customer's computer system experiences a failure, the FE is called and is responsible for diagnosing the problem and repairing the system. Certain maintenance contracts stipulate a certain timeframe in which the computer must be fixed, adding pressure to the FEs job. TTD provides training for both FEs and SAs.

While the FEs maintain the computer system hardware, the SAs are responsible for the system software. When the computer system fails, it is not always immediately known whether the problem is a hardware failure or a software failure. SAs and FEs often work together to determine the origin of the failure.

FEs and SAs report to DSMs. Typically 6 - 12 FEs and SAs report to one DSM, who works in a district office. The DSMs do the performance appraisals for the FEs and SAs, but while they periodically make site visits, they are rarely onsite during a computer

failure/repair. DSMs were chosen for the study partly because of the relationship between job performance appraisal and level 3 evaluation.

The survey will be sent to 165 recipients and includes FEs, SAs, and DSMs. They are personnel from the United States who work at "open" sites and secured sites, and personnel from international offices, also from both open and secured sites.

Instrumentation

A survey was used to gather data to aid in the development of a level 3 evaluation tool. The researcher worked with the manager of the training department, who is an expert in workplace technical training and documentation, to develop the survey questions. A Product Support Engineer was also consulted regarding sources of data that might be beneficial for training evaluation.

The researcher formed a list of effective level 3 evaluation methods based on the literature review. TTD needs to employ methods that can be easily implemented with the complex job duties and varying work environments of the student population. But it also needs methods that provide the most meaningful data. The list of evaluation methods provides the basis for the first two sets of survey questions.

A cover letter accompanies the survey and provides background information regarding the study, explains the different levels of training evaluation, and describes the purpose of the survey. The first set of questions asks the respondents to rate how much time they feel they can devote to each evaluation method. This information is important because TTD must regularly rely on input from training students and their bosses during the evaluation process and TTD does not wish to put a burden on these providers of valuable evaluation data. The second set of questions asks respondents to rate the value of the data from the evaluation methods, as it relates to how well newly learned skills are

applied on the job. A Lickert scale is used for the first 2 sections of the survey for ease of analysis.

The third section of the survey lists additional sources from which evaluation data may be available, and asks respondents to chose one that they feel provides the most valuable data. The three sources are:

- Performance improvement goals
- Customer satisfaction surveys
- Incident reporting database

Each employee of Company A develops performance goals as part of his or her annual performance appraisal. Employee records are confidential and TTD does not have access to employee performance *ratings*. However, TTD may be able to use the performance *goals* as means of identifying deficiencies in training and targeting needs for future training.

Customer satisfaction surveys are distributed annually to Company A customers. Company A takes very seriously the results from this survey and uses the results in many departments for quality improvements. TTD currently does not use customer satisfaction survey results. The customer satisfaction survey is comprehensive and covers everything from applications libraries to ease of contact with salespersons. It also includes a section on customer service, which serves as an indicator of satisfaction with the performance of FEs and SAs.

The incident reporting database is used by the FEs each time they respond to a customer call. The database records information such as response time, the type of failure, how it was resolved, and the time from the start of the incident to the time the system was returned to the customer in working order. This database was chosen as a possible source of evaluation data because while it cannot directly identify a training

issue, trends can be analyzed for potential evaluation use. For example, an analysis of incidents shows that memory failures are taking longer than anticipated to repair, or that certain memory parts are routinely replaced incorrectly. The training department can use this information and examine the memory troubleshooting section of the class and make improvements. Trends such as these may also indicate a need for follow-up or refresher training, particularly for subassemblies that have very low failure rates.

The survey also asks respondents if they know of any other sources of data that might be valuable in assessing how newly learned skills are applied on the job. The researcher asks respondents to chose only one of the data sources because the training department does not have the resources to spend a lot of time examining multiple sources of data, in addition to any survey or interview results, when performing level 3 evaluations.

Finally, the survey provides space for additional comments. Since this evaluation is new to the training department, as well as the rest of Company A, the training department is interested in any opinions respondents have regarding level 3 evaluation.

Survey Validation

The researcher reviewed the first draft of the survey with the training manager, who is an expert in technical training and documentation, a fellow course developer/instructor, and a product support engineer. Product support engineers provide the second line of support for computer system failures. If the FE or SA cannot resolve the problem, they escalate it to a product support engineer.

As a result of the survey review, changes were made to the scoring for first two sections of questions and the order in which the questions appear on the survey was also

changed. The training manager suggested adding definitions of the data gathering methods listed in the survey. The researcher also added a section that allows the respondents to identify their work environment (U.S., non-U.S., secured or non-secured site). This information helps TTD determine if they will need more than one method of evaluation distribution. The product support engineer provided an explanation of the types of information that is managed by the incident reporting database.

Survey Distribution

The survey will be emailed to the recipients. The email addresses are available via an employee database. The cover letter and a text version of the survey will be included in the body of the email and the survey will also be included as an HTML attachment for ease of responding. The survey must be sent in both text and HTML format because the study population uses many different computer platforms and different tools to read email.

Chapter 4

Results of Survey

Introduction

The purpose of this study was to research training evaluation methods and develop an evaluation plan that includes level 3 evaluations. The reason for researching this topic was because the Technical Training and Documentation (TTD) department at Company A currently uses only level 1 evaluations, which, when used alone do not provide comprehensive information regarding the quality and effectiveness of the training programs. This study is also important because the student population is geographically dispersed around the world, which makes it difficult to use traditional methods of post-training evaluation. The study also seeks uniform procedures for conducting evaluations because several of the instructors in the TTD department are contractors and work on a part time basis.

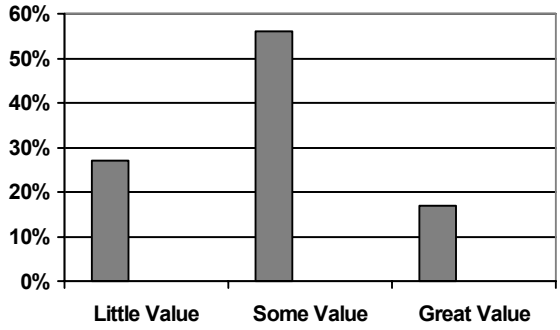
The survey was distributed by email in HTML and text format. The respondents were asked to return the survey within one week. After one week, 32% of surveys were returned. The researcher sent a follow-up email reminding respondents to complete the survey, and gave a one-week extension. After the second week, additional surveys were returned for a total of 46%.

Research Question One

The first survey question asked the respondents to rate the value of four evaluation methods in determining how well newly learned skills are applied to the job. Because the training department must rely on field engineers, site analysts, and district service managers for participation in level 3 evaluations, it is important to determine how valuable they feel the evaluation methods are. The majority of respondents, 56%, ranked

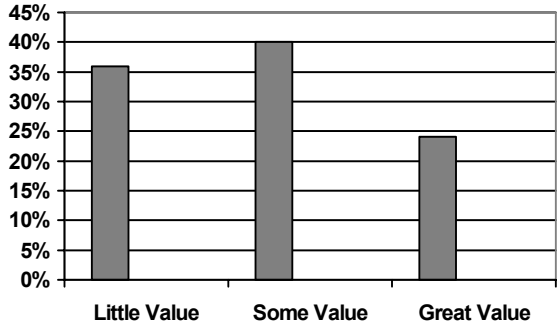
surveys as having some value. Twenty-six percent felt surveys have little value, and 8% felt they have great value. Figure 1 illustrates these results.

Figure 1. Value of Surveys



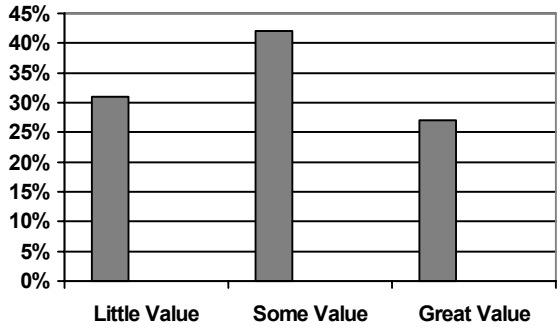
The responses for the interview method were more evenly spread, with 40% responding that interviews have some value. Thirty-six percent felt they have little value, and 24% rated interviews as having great value. Figure 2 illustrates these results.

Figure 2. Value of Interviews



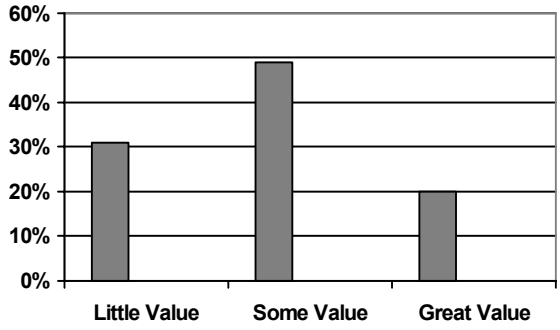
The responses for testing were fairly evenly distributed. Forty-two percent of respondents felt that testing has some value, 31% felt it has little value, and 27% felt it has great value. Figure 3 shows the distribution of responses.

Figure 3. Value of Testing



The last method listed in question one was follow-up calls. Forty-nine percent felt this method has some value, 31% rated it as having little value, and 20% rated it as having great value. Figure 4 shows these results.

Figure 4. Value of Follow-up Calls



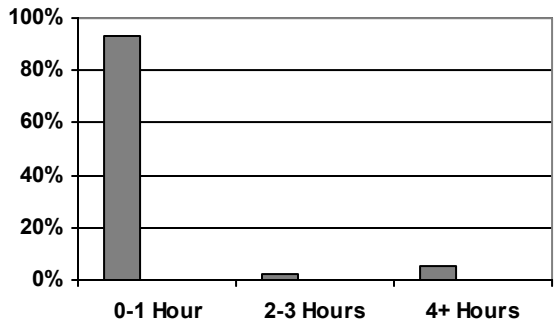
Summary of Question One

Most responses to question one were rather evenly spaced, except for the responses to the value of surveys as an evaluation method. This makes it difficult to determine that one method has substantially more value that any of the others.

Research Question Two

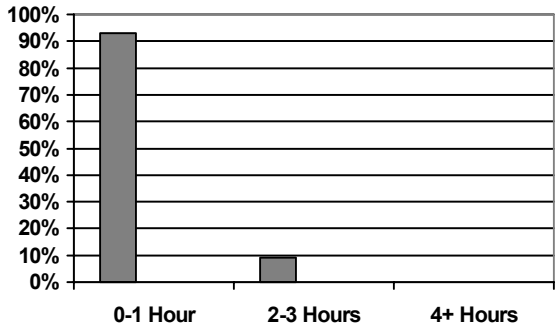
This survey question asked the amount of time the respondents thought it would take to complete the four evaluation methods. The training department does not want to burden the evaluation participants, so it is important to understand how much of a time commitment is required for the different level 3 evaluation methods. The responses for question two were much more polarized than those for question one. Ninety-three percent felt that surveys would require 0-1 hour of time to complete. Five percent thought surveys would take 4 or more hours and 2% felt they would take 2-3 hours. Figure 5 illustrates these results.

Figure 5. Time Commitment for Surveys



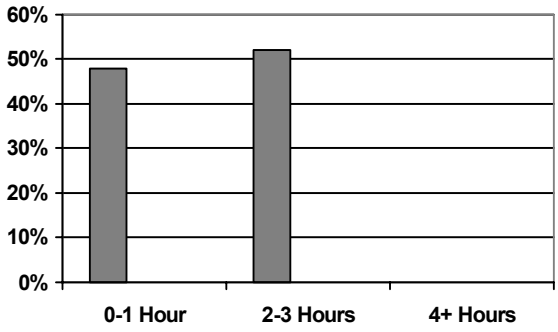
The responses for interviews indicated that 91% of respondents think that they will require 0-1 hour of their time. Nine percent feel interviews will take 2-3 hours and no respondents feel that they will take four or more hours to complete. Figure 6 shows these results.

Figure 6. Time Commitment for Interviews



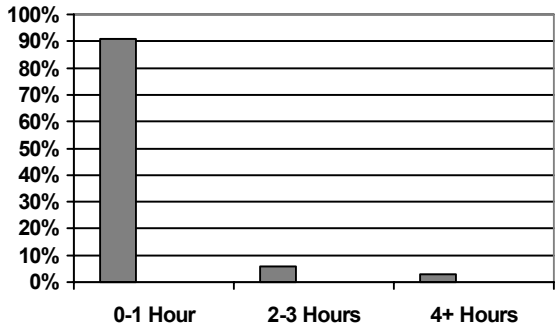
The responses for testing were fairly evenly spread across 0-1 hour and 2-3 hours. Fifty-two percent think testing will take 2-3 hours and 48% feel it will take 0-1 hour to complete. Figure 7 illustrates this data.

Figure 7. Time Commitment for Testing



When asked how much time they think it will take to complete a follow-up phone call, 91% feel it will take 0-1 hour, 6 % feel it will take 2-3 hours, and 3% think it will take 4 or more hours. Figure 8 shows this data.

Figure 8. Time Commitment for Follow-up Calls



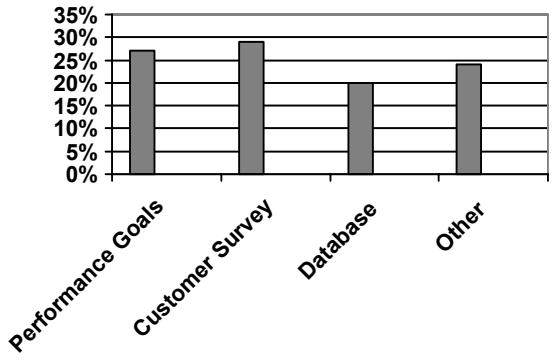
Summary of Question Two

The results for question two were more clear than for question one. Most respondents feel that the selected evaluation methods will take one hour or less to complete. This data is positive for the training department because participation in any training evaluation should not burden the participant.

Research Question Three

The third survey question asked the respondents to select one of the listed methods for gathering level 3 evaluation data. The three data sources were performance improvement goals, customer satisfaction surveys, and the incident reporting database. Respondents were also allowed to select 'other' and recommend their own source for level 3 evaluation data. Twenty-nine percent of respondents chose customer satisfaction surveys, 27% percent selected performance improvement goals, and 20% selected the incident reporting database, as shown in figure 9. Twenty-four percent recommended other data sources. Appendix D contains the list of other responses for question three. The data was entered exactly as it appeared on the survey. Many of the survey respondents use English as a second language, which may account for variances in grammar usage.

Figure 9. Other Data Sources



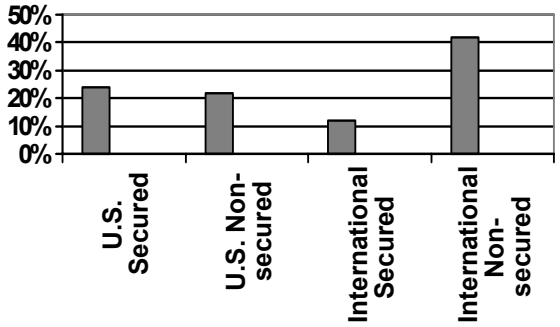
Summary for Question Three

Like survey question one, the results from question three were somewhat evenly spread among the four categories. But the additional data provided by the respondents will help the researcher in developing data sources for level 3 evaluations. Appendix D contains the written comments from question three.

Research Question Four

The last survey item asked the respondent to identify his or her work environment. The type of work environment is important because the training evaluator must be able to make contact with these individuals after the training event. Employees who work at special systems sites or in other countries can be more difficult to contact. Most respondents (42%) work outside the United States at non-special systems sites. Twenty-four percent work in the U.S. at special systems sites, 22% work in the U.S. at non-special systems sites, and 12% work outside the U.S. at special systems sites. This data is represented in figure 10.

Figure 10. Profile of Work Environments of Respondents



Summary for Question Four

The majority of respondents work outside the U.S., but at non-secure sites, and the next highest population (24%) work inside the U.S., but at secured sites. This can make contact difficult. Traditionally, the easiest employees to contact are those who work inside the U.S. at non-special systems sites, but only 22% of respondents work in that type of environment.

The end of the survey included a statement that invited respondents to make any additional comments. These comments are found in Appendix C. Many of the respondents use English as a second language and the comments were entered in Appendix C as they were written in the survey. This may account for variances in grammar useage.

Chapter 5

Summary and Results

Introduction

Corporate training departments provide many types of training programs for their employees. Sales, management, leadership, and technical training are just some of the training programs that are offered. Evaluation is an important step in the training process because it can provide a measure of the effectiveness of the training and identify areas for improvement.

The purpose of this study was to research different methods for performing evaluation, particularly those based on Kirkpatrick's four levels of evaluation. This study also sought to determine the most appropriate level 3 evaluation methods for the Technical Training and Documentation (TTD) department at Company A. Level 3 evaluations are usually conducted 3 - 6 months after the training event and their purpose is to measure how well the newly learned skills are applied on the job. The trainers in the TTD department face the additional challenge of trying to maintain contact with a student population that is located all around the world.

Methods and Procedures

After a thorough review of literature on the subject of training evaluation, the researcher derived a list of the most widely used methods for performing level 3 evaluations. These methods include surveys, interviews, testing, and follow-up calls. Research revealed that observations are also a very effective evaluation method, but because the student population is located around the world, it is not a feasible choice for TTD. The trainers in TTD must rely on participation from the students and their managers in the evaluation process. For this reason, TTD must employ an evaluation

process that is easy to use, does not put a burden on the participants, yet provides meaningful data. The researcher created a survey that covered four main subjects:

- Effectiveness of the evaluation methods
- Amount of time to complete each method
- Other data sources that might provide level 3 evaluation data
- Work environment of evaluation participants

It is important to understand how many participants work in secured work environments because those work sites have limited phone access and may or may not have email access. Geographic location is also an issue because phone calls may be difficult to use as an evaluation method when participants are located in different time zones.

The survey included definitions of surveys, interviews, testing, and follow-up calls so that the respondents had a clear understanding of each method.

Value of Methods

The first research question asked the value of the four evaluation methods. The results from this question did not indicate that any one method had greater value than another. The majority of responses indicated that each method has 'some value'. This was somewhat disappointing because it would be easier to select one of the methods if it had been shown to have greater value than the others. The researcher analyzed the data further by averaging the results for each method to determine whether each method had little or some value. For the survey method, 20 respondents felt they have little value. Forty-three chose some value and 13 chose great value. By averaging the responses for some value and great value, the result was 28. So 28 of respondents felt there is at least some value in surveys, compared to 20 who felt they have little value. From this data the researcher concludes that surveys are a valuable evaluation method.

The second method in research question one was interviews. By averaging the results for some value and great value, the data shows that 24.5 of respondents felt

interviews have at least some value, compared to 27 who felt they have little value. This indicates that interviews have little value.

The third method on the survey was testing. The averaging technique was used again, but with closer results. An average of 26 assigned at least some value to testing, compared to 24 who felt testing has little value.

The results for follow-up calls were identical to those for testing, with an average of 26 feeling that follow-up calls have at least some value and 24 who feel they have little value. Because the recipients of follow-up calls are located around the world in many different time zones, it would be difficult to conduct them consistently enough to use them as a meaningful evaluation method.

Time Requirement

The second question asked the amount of time needed to complete each of the four evaluation methods. The results for this question were more definitive. More than 90% of respondents felt that surveys, interviews, and follow-up calls would take 0 - 1 hour to complete. The results for testing were more evenly spread, with 48% choosing 0 - 1 hour and 52% choosing 2 - 3 hours. This information is beneficial because TTD does not want the evaluation method to be a burden on the participants. When an evaluation is easy to complete in a short amount of time, participants are more willing to comply. This data suggests that all of the methods are favorable with respect to the amount of time required to complete.

Other Data Sources

The third survey question asked the respondents to chose one of four data sources that could provide additional level 3 evaluation data. The data sources are:

- Performance improvement goals
- Customer satisfaction surveys
- Incident reporting database
- Other

Each employee of Company A develops performance goals as part of his or her annual performance appraisal. TTD may be able to use the performance goals as means of identifying deficiencies in training and targeting needs for future training.

Customer satisfaction surveys are distributed annually to all Company A customers. This survey includes a section on customer service, which serves as an indicator of satisfaction with the performance of FEs and SAs.

The incident reporting database is used by the FEs and SAs each time they respond to a customer call. The database records information such as response time, the type of failure, how it was resolved, and the time from the start of the incident to the time the system was returned to the customer in working order. As an example of how this data could be used by the training department, if the data shows a trend of memory failures taking a long time to resolve, the training department could examine the memory troubleshooting section of the class and make adjustments. The survey also asked respondents if they could suggest any other sources of data that might be valuable in assessing how well newly learned skills are applied on the job.

The results were rather evenly split among the four choices. This indicates that the respondents do not feel strongly about any one of the data sources. The TTD department could employ any of the data sources, and will probably choose one from which data can be easily retrieved.

Several respondents did provide other sources that may be used for level 3 evaluation data. Many of these comments suggested using the field personnel's contact with technical support as a means of evaluating how well skills are used on the job. Other responses called for more direct contact such as peer review, individual testimony, and managers' evaluation.

Work Environment

The last item on the survey asked the work environment of the population. The results were surprising because it was originally believed that students who work outside the U.S. or who work at secured sites are more difficult to contact by email. But the survey results proved that theory wrong. The highest percentage of survey respondents were from outside the U.S. and the second highest percentage of those responding were from secured work sites. This is beneficial for the TTD department because different work environments should not hinder efforts to make contact by email.

Recommendations

Of the four level 3 evaluation methods listed in the research questionnaire, surveys were found to have the most value. This supports the conclusion that a survey is an appropriate method for training evaluation for the TTD department. Both students and their managers should be included in the survey, with survey questions tailored for each group. The survey should be conducted 3-6 months after the training event and distributed by email.

The research also indicates that testing and follow-up calls have identical value as an evaluation tool. The use of testing after the training event can be difficult because the trainer may have little control over the test environment. However, the TTD department has a web site that has been specifically designed for its training audience. This web site

contains training schedules, registration information, as well as a full collection of online documentation and self study training programs. It also supports an online testing tool, which can be used to conduct level 3 evaluations. Specific tests will need to be developed to target skill application. The training department must email the link to the test on the TTD web site to the students at a predetermined time period after the training. When the student completes the test, the test tool automatically scores the test and distributes the results to the student and his or her manager, as well as training department personnel.

Although follow-up calls were ranked as being as valuable as testing, they are not a feasible evaluation method for TTD. When the students complete training, they return to their work sites, which are located all around the world. It would be impossible for TTD to make contact with students located in different time zones in a manner consistent enough to gather valid evaluation data.

The research also suggests the use of communications between field personnel and technical support as a source of level 3 evaluation data. The incident reporting database is the primary tool for logging the escalation of problems to the technical support level. The incident reporting tool can create reports that pull evaluation-specific information from the database, which can be emailed to TTD personnel. The reports can be set up for specific student groups 3-6 months after completion of training. The data in the reports can be analyzed for trends that may indicate deficiencies in the training.

Limitations of the Study

The way in which survey question one was structured limited the amount of information that was sought. The survey allowed respondents to assign a value to each evaluation method, which could have theoretically been identical for each method. By

having the respondents rank the four methods, or simply chose one of the four as the most valuable may have resulted in more definitive data.

Summary

The three main recommendations for the TTD department are that they employ surveys, online testing, and data gathered from the incident reporting database as their level 3 evaluation process. Surveys are to be distributed to students and their managers 3-6 months after the training program. Students will also be asked to complete a skills application online test from the TTD web site. Finally, TTD personnel will receive reports from the incident reporting database to track and identify field problems that may be linked to deficiencies or problems with the training programs.

References

Abernathy, Donna J. (1999, February). Thinking outside the evaluation box. Training & Development, 53, 18-23.

Bee, Frances. (2000, March). How to evaluate training. People Management, 6, 42-43.

Bernthal, Paul R. (1995, September). Evaluation that goes the distance. Training & Development, 49, 41-45.

Beveridge, Don. (2000, December). Training's ROI. Industrial Distribution, 89, 76.

Birkland, Carol. (2001, February). Making training work. Fleet Equipment, 27, 28-31.

Blanchard, P. Nick, Thacker, James W., Way, Sean A. (2000). Training evaluation: perspectives and evidence from Canada. International Journal of Training & Development, 4:4, 295-304.

Brinkerhoff, Robert O. (1989). Evaluating training programs in business and industry. San Francisco: Jossey-Bass.

Carnevale, Anthony P., Gainer, Leila J., Villet, Janice. (1990). Training in America. The organization and strategic role of training. San Francisco: Jossey-Bass.

Deming, Basil S. (1982). Evaluating job job-related training: a guide for training the trainer. Washington, DC: American Society for Training and Development; Englewood Cliffs, NJ: Prentice-Hall.

Dixon, Nancy M. (1996, May). New routes to evaluation. Training & Development, 50, 82-85.

Fisher, Sharon G. Ruffino, Barbara J. (1996). Establishing the value of training. Anherst, MA: HRD Press, Inc.

Friedman, Paul G., Yarbrough, Elaine A. (1985). Training strategies from start to finish. Englewood Cliffs, NJ: Prentice-Hall.

Goldwasser, Donna. (2001, January). Beyond ROI. Training, 82-90.

Haberstok, JoKay. (1997, June). Training feedback...it's what you do with it that counts. Journal for Quality & Participation, 20, 40-42.

Hawthorne, Elizabeth M. (1987). Evaluating employee training programs: a research-based guide for human resources managers. New York: Quorum Books.

Hubbard, Andrew. (2001, April). Training evaluation. Mortgage Banking, 115.

Keenan, William, Jr. (2000, June 12). Sales training ROI? Setting the proper objectives allows the most valuable evaluation. Industry Week, 249, 23.

Kidder, Pamela J., Rouiller, Janice Z. (1997, Spring). Evaluating the success of alarge scale training effort. National Productivity Review.

Kirkpatrick, Donald L. (1994). Evaluating training programs: the four levels. San Francisco, CA: Berrett-Koehler Publishers, Inc.

Laird, Dugan. (1978). Approaches to training and development. Reading, MA: Addison-Wesley Publishing Company.

Lawson, Karen. (1998). The trainer's handbook. San Francisco, CA: Jossey-Bass.

Lippert, Robert M., Radhakrishna, Rama, Plank, Owen, Mitchell, Charles C. (2001). Using different evaluation tools to assess a regional internet inservice training. International Journal of Instructional Media, 28, 237-248.

Mathews, Brian P., Ueno, Akiko, Kekale, Tuano, Repka, Mikkp, Pereira, Zulema Lopez, & Silva, Graca. (2001). Quality training: Needs and evaluation – findings from a European ssurvey. Total Quality Management, 12, 483-490.

May, Leslie Steven, Moore, Carol Ann, Zammit, Stephen J. (1989). Evaluating business and industry training. Boston, MA: Kluwer Academic Publishers.

Parry, Scott B. (1996, May). Measuring training's ROI. Training & Development, 50, 72-77.

Phillips, Jack J. (1996, February). ROI: The search for best practices. Training & Development, 50, 42-47.

Phillips, Jack J. (1996). Accountability in human resources management. Houston, TX: Gulf Publishing Company.

Phillips, Jack J. (1997). Handbook of training evaluation and measuring methods. Houston, TX: Gulf Publishing Company.

Phillips, Jack J. (1997). Return on investment. Houston, TX: Gulf Publishing Company.

State of the Industry. ASTD's Annual Review of Trends in Employer-Provided Training in the United States. February 2002.

Stufflebeam, Daniel L. (2001). Evaluation models. San Francisco, CA: Jossey-Bass.

Turnbull, Jeffrey, MD, Gray, Jean, MD, MacFadyen, John, MD. (1998). Improving in-training evaluation programs. Ottawa, Ontario, Canada: University of Ottawa, Department of Medicine.

Wade, Pamela A. (1994). Measuring the impact of training: a practical guide to calculating measurable results. Irvine, CA: Richard Chang Assoc.

Appendix A

Cover Letter

The Technical Training and Documentation (TTD) department plans to develop a more thorough training evaluation program. We are doing this because our goal is to deliver effective training that targets the job tasks of our students. Our current evaluation program consists of level 1 and level 2 evaluation methods.

- Level 1 measures student reaction to the training and is done with a questionnaire that is distributed to students at the end of the training class.
- Level 2 evaluation assesses the learning that takes place during the training. TTD utilizes quizzes, tests, and lab exercises to accomplish level 2 evaluation.
- Level 3 evaluations measure the extent to which the skills learned during training are applied on the job.

We do not currently use level 3 evaluations. Level 3 evaluations can help us improve our training courses by identifying deficiencies in the training and highlighting job tasks and skills that our training can target.

Level 3 evaluations are typically performed 3 – 6 months after the training class, so it is difficult for us to execute without relying on input from you. There are many different methods available for performing level 3 evaluations. We are conducting this survey to help us determine the most appropriate methods we can use. We will use the data from the survey to develop an effective level 3 evaluation program.

The survey asks your opinion regarding 4 different methods we are considering. Participation in level 3 evaluations involves a time commitment from you, but it will result in a better trained workforce, which benefits the entire company. Based on feedback from this survey, we will develop a process for conducting level 3 evaluations for our training courses. We will provide you with instructions to help guide you through your steps in the evaluation process.

Be assured that your participation in this survey is voluntary and will be totally confidential. If you chose to participate in this study, we request that you return your survey by April 19th so that we can begin processing the data. Thank you for your participation in this study.

If you have any questions regarding this survey, please contact:
Becky Pearl
Email address
Phone number

Appendix B

Survey

You may complete the survey by either replying to sender, or by printing a hardcopy and mailing it to _____. Please respond by April 19th so that we may begin processing the data. Definitions of terms are included for clarification.

1. For this part of the survey, please rate the value of the methods in determining how well newly learned skills are applied to the job.

	1 – Little Value	2 – Some Value	3 – Great Value
Surveys			
Interviews			
Testing			
Follow-up Calls			

2. Please rate the following methods based on how much time you think it would take to complete.

	0 – 1 Hour	2 – 3 Hours	4 + Hours
Surveys			
Interviews			
Testing			
Follow-up Calls			

Survey: A question and answer format administered by email. The questions would solicit factual data and/or opinions.

Interview: A telephone conversation with structured questions regarding how skills are being applied on the job.

Testing: A written question/answer instrument that would be distributed to students at some time period after training.

Follow-up Calls: Phone calls that gather specific data regarding the training class and skills that were taught during the training.

3. We are also considering gathering data from other sources that may provide insight to training-related work performance. Please select one of the following sources of data that you believe offers the most meaningful information regarding how skills have improved.

Performance improvement goals	
Customer satisfaction surveys	
CRUISE database	
Other (please specify)	

4. Please check the box that most closely defines your work environment. This information helps us to determine the best method to distribute an evaluation instrument and identify any special requirements for different work environments.

U.S. non-special systems	
U.S. special systems	
International non-special systems	
International special systems	

If you have additional comments or questions, please write them in here. Let us know if we can provide you with additional information regarding level 3 training evaluation.

Appendix C

Written Comments from Survey

“I think training manuals should have a lot of examples actually used on jobs. (Company A) manuals just provide us command’s syntax and meaning. I want to know how to use it.”

“I always believed that tests are the best tool. You have to trust the person to be honest when taking the test.”

“I think a survey is best. This would identify what parts of the class had a direct impact on our customer systems and the parts that did not. Testing is definitely the worst. Many time 3-6 months after taking a class much of what I have learned is a hazy memory if I haven’t had the opportunity to use the training. I.e., how many can successfully bust their first dump 6 months after taking the kernel debugging class?”

“My arrival into the (Company A) training world happened 4 days after joining - so A LOT WAS NEW and all references to "other systems or the way it was done previously on xyz " meant nothing. The training had too much emphasis on theory without the " links to the overall impact " of how it all was interconnected and affected different parts of the system.
And although the hands on exercises were interesting - it was not enough (I spent lot of my time down in the basement after hours and weekends to " play with the system " and that was most beneficial.) “

“What really gave me the " nuts and bolts" experience was the time after the training when I had the opportunity to spend some time on the factory floor working with people. Although initially I was trailing an engineer and learning - but later working by myself and doing testing on repaired modules and completing test runs.”

“Customers want systems running all the time so requesting a PM session for the sole purpose of doing testing and trying out redundant system combinations - in case there is a failure in not too distant future- yes it can be done but not too frequent.”

“The online documentation is great value. I regularly (2-3 times a month) look for some new information. Sometimes it can take time to find the info – but it is there. Some real ‘old timers’ they know exactly where to look. Hope my thoughts are of some value – big machines – very complex – seldom fails – delivery of leading edge training methods = 64$Q.”

“I notice from the mailing list for this email is all people in the Asia Pacific region. The two points I would raise are that interviews over the telephone to Australia lasting an hour will be very expensive and that you will probably have to get up in the middle of the night to conduct them as we are 15 hours ahead of you. I think that when creating training surveys you should take into account that there are really two types of engineers that you

teach. The first are new hires. These guys will know very little about (Company A), but they should have an EIC (engineer in charge) back home who will take them under their wing once they return from training. All the interviews and testing should be carried out by the EIC and he can pass back to you his thoughts on what his guys know or don't know after the course. The second type of student in the 'old timer' type who have been exposed to (Company A) gear a long time and think we can fix anything!! The old timers can submit a written report about how they have handled repair of the system after the training course as will as make suggestions for things that should be included or even excluded from future courses."

"I think we need to address on-going training needs. Often product lines evolve (Products X, Y) but return trips to Chippewa for each evolution could not be justified. Some method of providing additional training would be useful either to teach changes or just as a refresher. This could be CBT base, or perhaps even dedicated sessions booked on the training systems in CF run remotely to practice diagnostic techniques and troubleshooting skills. Also on the continuation training theme, I believe that one way to 'top-up' knowledge in the field is to have the HPS/SPS people provide short 'master classes' based on their wealth of practical experience on their particular product. As HPS folks in particular tend to visit all geographies occasionally this is something that could be achieved at minimum cost. This would be a good way to augment the mainly theoretical training that we receive from the training department."

"Based on (Company A's) technical service workforce, all with many years of experience, I believe we all know learning to understand function and repair processes of new equipments is fundamental to our mission. I also believe that there are no young job hoppers in (Company A) with no or little relations to the company. Only those need to be found out with tests to eliminate them."

"An evaluation 3-6 months after training is obviously going to be of help, but how about a second evaluation (perhaps on a sample of students) 12 months later? This would give people more time to come across problems and hence realize what would have been useful to have learned in training."

"Surveys would be easiest to complete if sent out in text format or posted on the web with a GUI set so that a clock answers the questions. This particular email had the attachment that could not be read with my UNIX mail tool, so I forwarded same to a Yahoo account and accessed this with Netscape, which displayed the survey in html but would not print same. Sooo I downloaded the attachment and displayed it but it did not want to let me enter anything in the blocks…..I could have worked through this but opted to copy and paste into file and with minor rework generate it this way."

"With the current product range except the (Product X) there are so few failures that I find I need to refer to my documentation nearly every time something comes up. The handiest thing to have from training is good references on 'where to find it'."

"It is very apparent from being in the field that our system failure rate is extremely low (except for one system) - but this means the field is getting " rusty " very quickly after graduating from the training."

“The field (my view) spends a substantial amount of time - investigating " potential " problems - how to fix these potential problems – planning upgrades - reading documentation - even so after a period of doing this - the "common " occurring problems become very easy to fix. You can read so much about our different products - but if the knowledge is never used, it will fade into the low lit areas of the brain.”

“Discussion here on-site suggested that perhaps the skill necessary to maintain a particular product is not taught in a classroom, but rather comes from experience at having faced similar problems on other products.

This is not saying that classroom training is not necessary, quite the contrary, the foundation for understanding the new product is gained in the classroom, but the actual skill necessary to make the repair, although demonstrated in a class or lab environment, can only be perfected through repetition. As an example, the soldering techniques necessary to repair (Product X) and (Product Y) modules, was demonstrated in a class environment, but most of us perfected our technique through repetition in the field. With (Product X) systems we were no longer required to repair modules at the component level, but the new techniques we had to learn were more in the area of software development and diagnostic manipulation. We were forced to learn how to operate the Maintenance Work Station when few of us had any previous operator and operating system exposure. The actual working of the mainframe computer changed little all the way up through
(Product X). Other aspects of the computers changed greatly...module sizes, connectivity between modules, channels and channel cabling vs. (Product X) systems, but the concept of getting data and instructions from point A to point B never changed.

The cooling systems on (Company A) equipment have been one of the aspects of the big computers that has probably changed the most. Yet, there has been little in the way of mechanical training and it is this field that still poses the biggest problem with the largest number of field folks. Although all of my people have received the required classroom exposure to the mechanical training, we still have only a couple of "experts" at the site level and most if not all repairs fall to these few. Does that mean that I should grade or evaluate everyone the same? Different people have different levels of expertise at various tasks. Most can do it all in one form or another, but the best person for the "job or task" at hand should be the individual who will have the greatest chance of making the repair in the shortest amount of time. Others should be involved, but this is not always possible when you live in a multi-system environment.

The training department needs to be made aware by engineering, development, manufacturing and checkout what is going to be required to repair the product. The training department needs to develop training that will touch on every aspect of what will be required for repair and define the techniques used in all the departments to teardown, repair and assemble the product in the most effective and efficient manner. These techniques should be flexible and changeable so that if an individual figures out a better method we can immediately incorporate the new method.

A suggested avenue of information could be the global support organization. If the training department evaluates communications between the field and support for new methods and techniques, or develop new methods from what is being sought from support as advise or expertise.

Following up on a training cycle could be fruitless if the individual had not been called upon to demonstrate the skills learned in the classroom. The answers received back could be quite ambiguous. (Ex. Question: Were the soldering skills learned in the classroom environment sufficient to allow you to perform on-site module repair? Answer: Yes. Question: How many module have you repaired? Answer: Well, none yet!) If the follow-up question is not asked, "How many modules have you repaired?" Then, the answer to the first question may not be true. Then if you find that this same individual, or other individuals are requesting help from support determining pin locations, and wire placement, etc. for module repair, then one could ascertain the deficiencies with the training for module repair. (They can fix the modules, but can't find the pin locations.) EQUALS, more intense training on pin and wire locations and how to find them. Even after they think they know how it all works. Support seems like an excellent source of information for training. I hope I haven't bored you too much with my ramblings."

"How to keep the support people and their skills finely tuned: Have a lot of systems with lots of problems - BAD FOR BUSINESS - OUT OF BUSINESS.

Repetitive training with more emphasis on " real system environment "

Cross train H/W with more S/W experience - system admin. - this has to be initially achieved outside a "real" site.

Then again - any new skill has to be maintained at the right momentum. Return visit to training center ----- Physical visit too expensive at this point in our sales cycle. Remote training - web - quiz. Regarding installations or working in CF - unfortunately expensive.

Electronic info sheets - "How to FIX RAID systems" - already available today.
Note: This was in a book prepared by ? Johnson - I can find it if you want - Great value.

I must say our self training programs we have on line - are of very high standard. If not too expensive to develop - do regular updates to enhance the program. Must ensure the Qs remain applicable to real problems - not just a reading skills."

"Very little emphasis has been put on system-level training: interaction between mainframe and I/O, different nodes on a ring, etc. For example: on a (Product X), after running (Test A) you need to run (Test B) before deadstarting the mainframe or you get MMR errors. Another example: on a (Product Y), after running (Test B) you sometimes need to reset (Product Z) or even reboot the workstation. These and more are not taught, they are learned, usually at 3:00 am and are horribly documented, or by trial and error."

Appendix D

Other Data Sources from Research Question Three

Peer reviews

Monitor emails/phone calls to tech support

Individual testimony

Manager evaluation

Hardware product support

None of these can be applied to solicit this information

Instructors should monitor the HPS email from the field to see what is giving the field the most problems

Ask the subject's manager if the training has been effective

Figure out how long and how successful folks are fixing problems by looking at MTTR

I would think performance evaluation rather than performance goals would be a better measure.